J Mol Evol (2016) 82:27–37
DOI 10.1007/s00239-015-9725-5

ORIGINAL ARTICLE

Evolution of Nine Microsatellite Loci in the Fungus *Fusarium oxysporum*

Jill E. Demers[1,2] · María del Mar Jiménez-Gasco[1]

Received: 23 December 2014 / Accepted: 19 November 2015 / Published online: 10 December 2015

Abstract The evolution of nine microsatellites and one minisatellite was investigated in the fungus *Fusarium oxysporum* and sister taxa *Fusarium redolens* and *Fusarium verticillioides*. Compared to other organisms, fungi have been reported to contain fewer and less polymorphic microsatellites. Mutational patterns over evolutionary time were studied for these ten loci by mapping changes in core repeat numbers onto a phylogeny based on the sequence of the conserved translation elongation factor 1-α gene. The patterns of microsatellite formation, expansion, and interruption by base substitutions were followed across the phylogeny, showing that these loci are evolving in a manner similar to that of microsatellites in other eukaryotes. Most mutations could be fit to a stepwise mutation model, but a few appear to have involved multiple repeat units. No evidence of gene conversion was seen at the minisatellite locus, which may also be mutating by replication slippage. Some homoplastic numbers of repeat units were observed for these loci, and polymorphisms in the regions flanking the microsatellites may provide better genetic markers for population genetics studies of these species.

Keywords Microsatellite · Minisatellite · Phylogenetics · Asexual fungi · Population genetics

Introduction

Microsatellites, or simple sequence repeats, are tandemly repeated regions of 1–6 DNA base pairs. Microsatellites are widely used as genetic markers, particularly in population genetics, as their high mutation rate makes them more informative than non-repetitive sequences when comparing closely related organisms. Understanding how microsatellites evolve has important implications for their use as molecular markers. Assumptions made in mutational models can affect population genetic analyses such as measures of diversity, population structure, population size, and evolutionary history (Putman and Carbone 2014 and references therein), as well as timing of molecular clocks (Sun et al. 2012).

Microsatellites are theorized to mutate primarily by replication slippage, in which misalignment of DNA strands during replication leads to the addition or subtraction of repeats (Levinson and Gutman 1987). Mutation due to replication slippage occurs at a much higher rate than that of point mutations, at a minimum of 10^{-4} mutations per division in the yeast *Saccharomyces cerevisae* (Henderson and Petes 1992). Minisatellites, consisting of 10–15 bp tandem repeats, are believed to mutate both through replication slippage and through recombination, especially gene conversion (Richard et al. 2008). Microsatellites may also mutate through homologous recombination, but this mechanism seems to be less

Electronic supplementary material The online version of this article (doi:10.1007/s00239-015-9725-5) contains supplementary material, which is available to authorized users.

✉ María del Mar Jiménez-Gasco
jimenez-gasco@psu.edu

Jill E. Demers
jill.demers@ars.usda.gov

[1] Department of Plant Pathology & Environmental Microbiology, The Pennsylvania State University, University Park, PA, USA

[2] Present Address: USDA-ARS Systematic Mycology and Microbiology Laboratory, 10300 Baltimore Ave., Beltsville, MD 20705, USA

common for microsatellites than for minisatellites (Richard et al. 2008). Microsatellite evolution has been described as a life cycle, with microsatellites being "born" by the formation of repetitive sequences, then expanding and contracting during their "life," and finally "dying" when the repetitive sequence degenerates (Kelkar et al. 2011). Microsatellites may originate from cryptically simple sequences, regions that contain repetitive patterns but less than 8 nucleotides of perfect repeats (Chambers and MacAvoy 2000), either through mutations in the cryptically simple sequence or from insertion events creating a duplicated sequence (Buschiazzo and Gemmell 2006; Kelkar et al. 2011). After formation of the microsatellite, mutations can add or subtract units, generally one unit at a time, although larger changes can occur (Henderson and Petes 1992). Point mutations may occur within repeats, interrupting the repetitive sequence and slowing the rate of mutation (Ananda et al. 2014). Accumulation of these interruptions is the primary cause of "death" of microsatellites (Kelkar et al. 2011). Buschiazzo and Gemmell (2006) suggest that the cryptically simple sequence left from a "dead" microsatellite could undergo a renaissance and become a new microsatellite.

Although microsatellites have been extensively studied in a number of organisms, less attention has been paid to microsatellites in fungi (Dutech et al. 2007), with the exception of *S. cerevisae* that has acted as a model organism in many important microsatellite studies (Richard et al. 2008). Microsatellites in fungi tend to have fewer repeat numbers and be less polymorphic than microsatellites in other eukaryotes (Dutech et al. 2007; Karaoglu et al. 2005; Schoebel et al. 2013). Microsatellites with fewer repeat units have a lower mutation rate than loci with more repeat units due to the instability of longer repetitive regions (Brandström and Ellegren 2008; Kelkar et al. 2008), but the reason why fungi tend to have shorter microsatellites is unclear. Fungi also tend to have lower microsatellite densities than most eukaryotes (Katti et al. 2001; Tóth et al. 2000; but see also Dieringer and Schlötterer 2003), most likely due to the low density of transposable elements in fungal genomes [e.g., 4 % of the *Fusarium oxysporum* genome (Ma et al. 2010) compared to 45 % of the human genome (International Human Genome Sequencing Consortium 2001)], as microsatellites are associated with transposons in many organisms, including humans (Kelkar et al. 2011) and fungi (Labbé et al. 2011). The low density of microsatellites in general may explain the low numbers of microsatellites with many repeats. Microsatellites in fungi could possibly be affected by fungal life cycles, as many fungi are predominantly haploid and asexual with occasional diploid, sexual stages, and mutations of microsatellites may occur during meiosis (Richard et al. 2008); however, the lack of sexual recombination does not appear to affect microsatellite evolution (Kayser et al. 2000; Paun and Hörandl 2006; Weetman et al. 2006).

Because of the challenges with analyzing microsatellite data, many authors have begun using single nucleotide polymorphisms (SNPs) found in the regions flanking microsatellites as additional genetic markers (Barthe et al. 2012; Payseur and Jing 2009). Regions directly adjacent to microsatellites are believed to be highly polymorphic as a consequence of the microsatellite mutational process (Brohede and Ellegren 1999; Santibáñez-Koref et al. 2001). These regions are easier to locate than other polymorphic regions because they can be found along with microsatellites through analysis of genomic sequence or throughout enrichment of microsatellite regions in genomic libraries. This approach has been used in several fungal phylogenetic studies (for example, Dettman and Taylor 2004; Liu et al. 2014).

As microsatellites are often used as molecular markers for the identification and study of fungi (Dutech et al. 2007; Taylor and Fisher 2003), having an accurate model of microsatellite evolution in fungi is important. In this study, the patterns of mutations of nine microsatellite loci and one minisatellite locus were investigated using a phylogenetic approach in three haploid species in the genus *Fusarium*, in considerable phylogenetic depth across 41 fungal isolates. The goal of this study was to assess if differences exist in the microsatellite life cycle in fungi which could account for the fewer repeat units and less polymorphic nature of microsatellite in fungi. A second goal was to determine the suitability of SNPs flanking microsatellites for phylogenetic reconstruction.

Materials and Methods

Fungal Isolates and Sequenced Genomes

The microsatellite loci were amplified for 3 isolates of *Fusarium redolens* and for 27 isolates of *F. oxysporum* (Supplementary Table 1), chosen as representatives of the known diversity of this species based on previous phylogenetic studies (O'Donnell et al. 1998; Demers and Jimenez-Gasco unpublished data). Sequences were obtained for an additional ten isolates of *F. oxysporum* from publically available genome sequences (Fusarium Comparative Database 2012; http://www.broadinstitute.org), as well as for the taxon used as outgroup, the sexually reproducing fungus *Fusarium verticillioides* (Supplementary Table 1). Sequences from the isolate *F. oxysporum* NRRL 26406 were obtained by sequencing each locus and not from the sequenced genome, which was released after the initial experiments.

Microsatellite Loci and Gene Markers

Primers for nine microsatellite loci (FoAB11, FoAD12, FoAG11, FoDC5, FoDE7, FoDF7, FoFA4, and FoFE5) and one minisatellite locus (FoED2) were developed in a previous study (Demers et al. 2014). These markers were developed for population genetics analysis of *F. oxysporum*. Primers were designed from a genomic library enriched for microsatellite repeats, created from an isolate of *F. oxysporum*. Only microsatellites identified as polymorphic within *F. oxysporum* were included in this study. DNA extraction, PCR, and DNA sequencing of the loci were done as previously described (Demers et al. 2014). The number of repeats in each microsatellite sequence was determined by sequencing and manual counting, and the fragment size was determined from the sequence data. Microsatellite data were mapped onto a phylogeny based on the translation elongation 1-α gene (TEF), a conserved house-keeping gene commonly used for phylogenetics of *Fusarium* species (O'Donnell et al. 1998). TEF sequences were obtained from GenBank, sequenced genomes, or by amplification and sequencing with the primers EF1 and EF2 according to previously published conditions (O'Donnell et al. 1998). All sequences were initially aligned using the program SEQUENCHER v. 4.1.4 (Gene Codes Corp.) and then manually adjusted.

Data Analysis

All analyses of microsatellite diversity and mutational model were done on the *F. oxysporum* isolates, excluding *F. redolens*, *F. verticillioides*, and any *F. oxysporum* isolate with more than one copy of that microsatellite locus. Analyses were based on the number of repeats or the size of the repetitive region, not on amplicon size to avoid homoplasy from indels in the flanking regions.

Testing of the stepwise mutational model versus the multi-step mutational model was done using a maximum likelihood approach with the program MISAT, using the default program settings (Nielsen 1997). Microsatellite gene diversity was calculated based on $1 - \sum x_i^2$, where x_i is the frequency of the ith allele (Nei 1987). The average pair-wise differences per base pair were estimated for the flanking sequences, including both putatively coding and non-coding bases but not insertions and deletions using the program SITES (Hey and Wakeley 1997). Alignment and comparisons of the genomes of *F. oxysporum* Fo47 and *F. verticillioides* 7600 were done using the program Mauve with default settings (Darling et al. 2004). Phylogenies for TEF and for the regions flanking the microsatellites were inferred using MrBayes 3.2.2 (Ronquist and Huelsenbeck 2003). The Bayesian phylogenetic analysis used the GTR + Γ + I model and was carried out for 1,000,000 generations of six MCMC chains, with the first 25 % of generations discarded and tree sampling every 200 generations.

In mapping the microsatellite sequences to the TEF phylogeny, microsatellite sequences from *F. redolens* were generally presumed to represent a more ancestral form of the microsatellite, as *F. redolens* appears to be basal to *F. oxysporum* and *F. verticillioides* (Skovgaard et al. 2003). However, sequences in *F. redolens* have also been evolving since these species differentiated, complicating the determination of ancestral versus derived characters. Because replication slippage occurs at a faster rate than mutations in non-repetitive sequences (Henderson and Petes 1992), *F. redolens* sequences containing long repeat arrays were also assumed to have mutated since the time of speciation. Incorrect determination of the ancestral state should not affect results beyond assuming the addition of repeat units when repeat units had been lost and vice versa. The most parsimonious series of mutations to create the repetitive sequences in *F. oxysporum* from the cryptically simple sequences in *F. redolens* and *F. verticillioides* was assumed.

Results

Sequencing of the Microsatellite and Minisatellite Loci

Sequences for the nine microsatellite loci were obtained for 37 isolates of *F. oxysporum*, three isolates of *F. redolens*, and one isolate of *F. verticillioides* (Supplementary Table 1), except for locus FoAG11, which did not amplify for the *F. redolens* isolates, and locus FoDD7, which was not found in the genome sequence of isolate *F. o.* PHW815. Each microsatellite locus was present in one copy in the eleven available *F. oxysporum* genomes and the *F. verticillioides* genome (http://www.broadinstitute.org), except for FoDC5, which was present twice in the genomes of isolates *F. o.* MN25, *F. o.* cotton, and *F. o.* PHW815, and FoDD7, which was present twice in the genome of isolate *F. o.* HDV247, although these may be artifacts of the genome assembly and not true duplications.

The minisatellite FoED2 was amplified for 23 of the 27 isolates of *F. oxysporum* and none of the *F. redolens* or *F. verticillioides* isolates. Non-homologous regions not containing the minisatellite were amplified for *F. oxysporum* isolates MX4 and ET5A in multiple attempts. The minisatellite was not found in six of the eleven sequenced genomes, including isolate *F. o.* NRRL 26406, from which it was amplified by PCR from extracted DNA. All genome sequences that contained FoED2 had different repeat sequences than those individually sequenced, suggesting

that the genome was not well assembled at this locus, and all genome data were excluded from further analysis. Based on genome annotation, FoED2 is located in a putative transposable element (http://www.broadinstitute.org), which may explain why non-homologous regions were amplified by PCR or the locus was not found in some isolates.

Evolution of the Nine Microsatellites Across the *Fusarium oxysporum* Phylogeny

The microsatellite loci were mapped onto a tree of the translation elongation 1-α gene (TEF) to analyze the evolution of these repeats (Fig. 1). The microsatellites were conserved across the three species and were identified in the outgroup *F. verticillioides*. Alleles with similar number of repeats and similar interruptions of the microsatellites were phylogenetically related as expected. However, nearly all nine loci, with the exception of FoFE5, had instances where isolates in different parts of the phylogeny had the same number of microsatellite repeats (Fig. 1), most likely due to homoplasy, although they may also be due to conservation of an ancestral sequence in some branches of the phylogeny. Amplicon sizes were even more likely than repeat number to be shared among unrelated taxa due to insertions and deletions in the flanking regions, such as for FoAB11, FoDC5, and FoDD7. Size homoplasy was also observed with the compound microsatellite FoAG11, in which differing numbers of each repeat unit could produce the same size fragment, for example, $(CT)_6(GT)_{10}$ and $(CT)_8(GT)_8$.

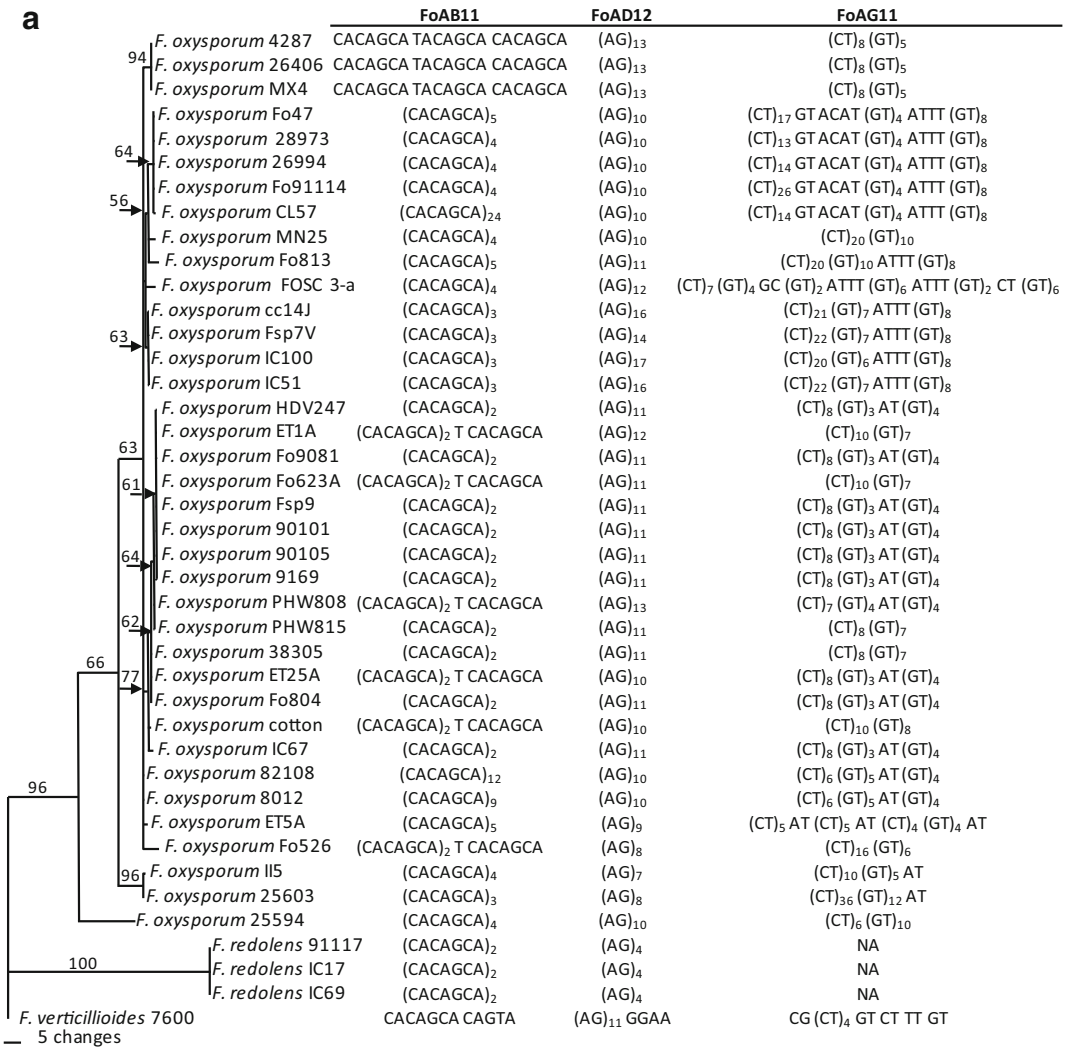

Fig. 1 Sequences of the nine microsatellite loci mapped onto a phylogenetic tree of all isolates based on the translation elongation factor 1-α gene (maximum parsimony analysis, 1 of 75 most parsimonious trees, *F. verticillioides* 7600 set as an outgroup). Bootstrap values >50 % given for each branch (1000 replicates). Sequences marked with an *asterisk* had two identical copies per genome. *NA* no amplification or locus was not present in sequenced genome or two different copies were present per genome

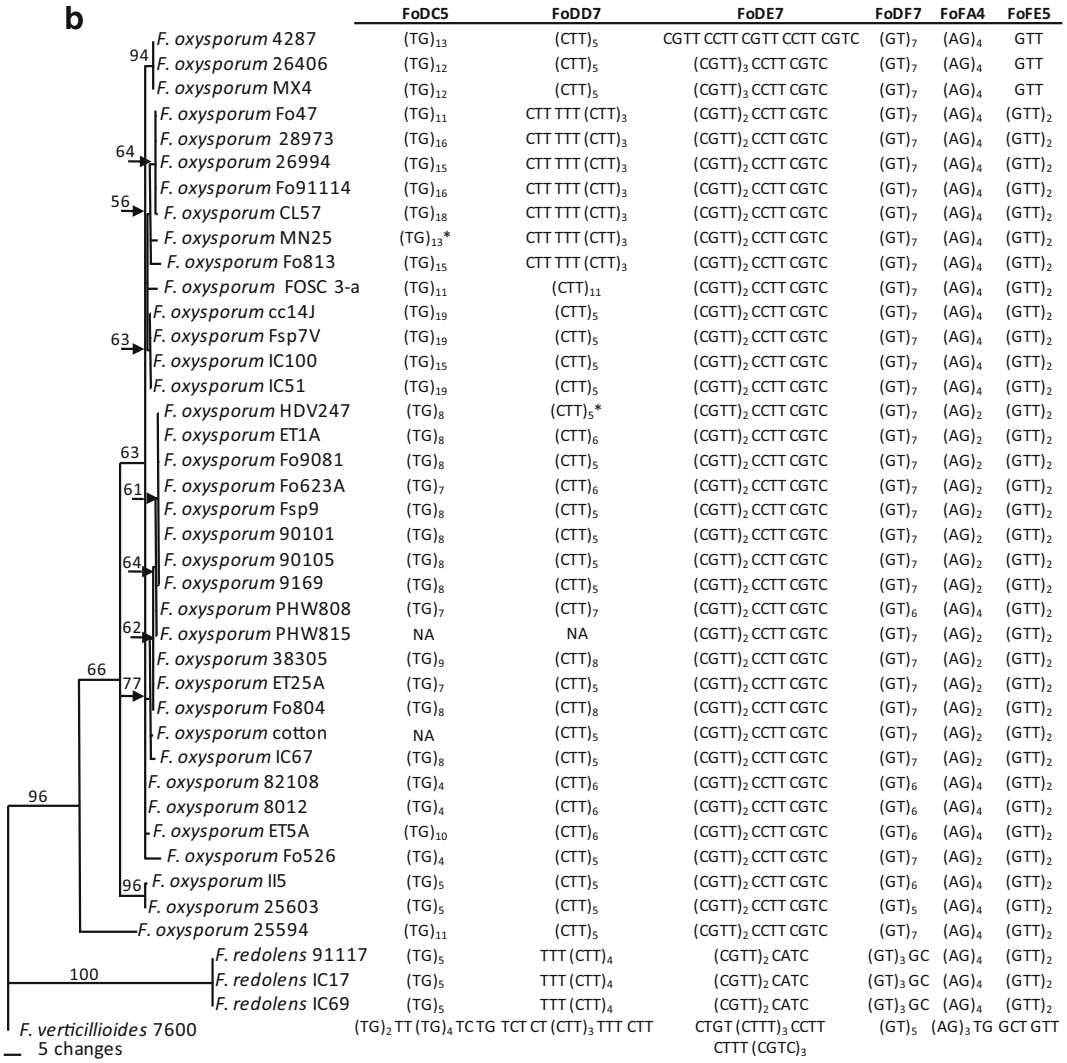

Fig. 1 continued

The minisatellite FoED2 was also mapped onto a TEF gene phylogeny (Fig. 2). This minisatellite had four variant repeat units (coded A through D in Fig. 2). Only one of the units (coded B) was polymorphic in number of repeats across *F. oxysporum.* No rearrangement of repeat units was observed.

Stepwise Versus Multi-step Mutations

Loci with larger repeat arrays, such as FoAB11, FoAD12, FoAG11, and FoDC5, were the most diverse, and loci with more allele diversity also frequently had higher rates of polymorphisms in the flanking regions (Table 1). The two least diverse microsatellites were FoDE7, located in the 3′UTR of a hypothetical protein, and FoFE5, located in the exon of a NAD-specific glutamate dehydrogenase gene. No relationship was observed between motif size and diversity (Table 1).

The data were fit to two mutational models, a stepwise model assuming that only one repeat unit is gained or lost during replication slippage and a multi-step model that allows for mutations involving more than one repeat unit. The multi-step model was inferred to be a significantly better fit than the stepwise model for loci FoAB11, FoAG11, and FoDD7, while the stepwise model was significantly better for loci FoAD12 and FoDC5 (log likelihood ratio test, $P < 0.05$) (Table 1). Possible multi-step mutations can be seen in the sequences mapped to the TEF phylogenetic tree for locus FoAB11, in which isolate *F. o.* CL57 has 24 repeats compared to 4 repeats in close relatives, and for locus FoAG11, in which isolate *F. o.* Fo91114 has 26 CT repeats compared to 13 or 14 repeats in close relatives (Fig. 1). Non-adjacent duplications of interruptions of the repeat region can also be taken as evidence of multi-step mutations, as the same interruption is unlikely to occur twice (Dettman and Taylor

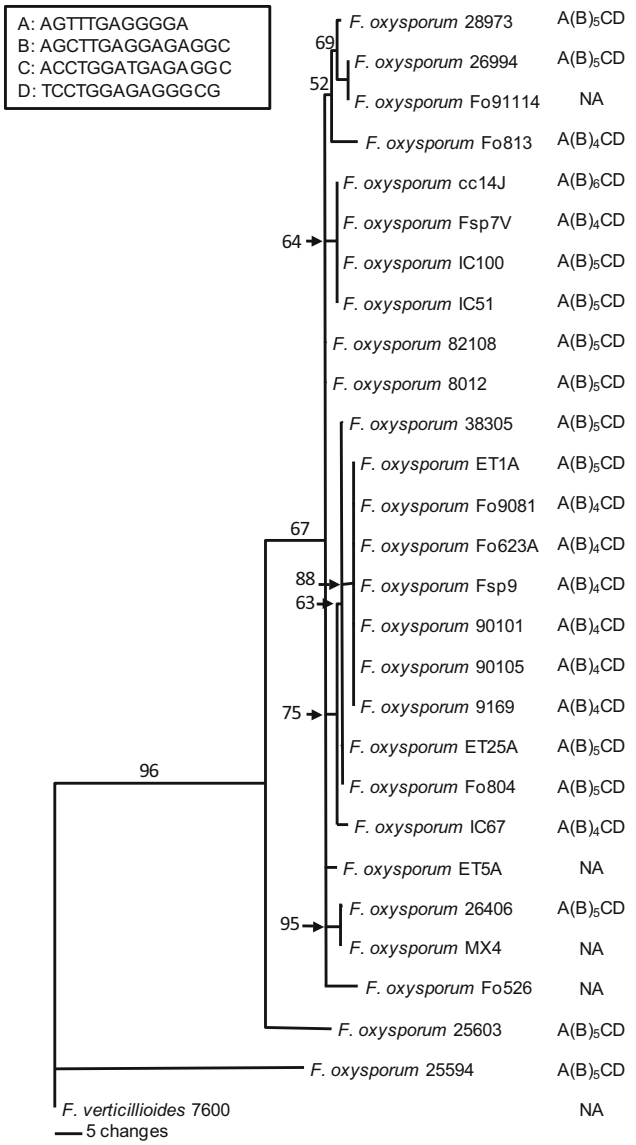

Fig. 2 Sequences of the minisatellite FoED2 mapped onto a phylogenetic tree based on the translation elongation factor 1-α gene (maximum parsimony analysis, 1 of 6 most parsimonious trees, *F. verticillioides* 7600 set as an outgroup). Bootstrap values >50 % given for each branch (1000 replicates). *NA* no amplification or a non-homologous sequence was amplified. Letter codes for each variant repeat are given in the *upper right*

2004). In FoAG11, the interruption ATTT was observed twice in isolate *F. o.* FOSC 3-a, suggesting a slippage event involving up to 9 repeat units, while the sequence $(CT)_5AT$ occurs twice in isolate *F. o.* ET5A, suggesting a slippage event of up to 6 repeat units (Fig. 1).

Formation of the Nine Microsatellite Loci

The microsatellite sequences were mapped to the TEF gene tree (Fig. 1) and studied to determine if the most parsimonious explanation for the formation of repetitive sequences is base substitutions in cryptically simple sequences or short sequence duplications. The compound microsatellites FoAG11 and FoDE7 appear to have been formed by base substitutions in the cryptically simple sequence surrounding the microsatellite. For locus FoAG11, *F. verticillioides* contains a pure microsatellite surrounded by cryptically simple sequence: CG $(CT)_4$ GT CT TT GT, while *F. oxysporum* contains the perfect compound microsatellite $(CT)_n(GT)_n$. The $(GT)_n$ repeat in *F. oxysporum* may be derived from an ancestral version of the cryptically simple sequence in *F. verticillioides*. For locus FoDE7, the pure microsatellite $(CGTT)_n$ is followed by the cryptically simple sequence CCTT CGTC in *F. oxysporum* and by CATC in *F. redolens*. In *F. verticillioides*, two different repeat motifs are repeated in the complex microsatellite CTGT $(CTTT)_3$ CCTT CTTT $(CGTC)_3$. The microsatellite in *F. verticillioides* therefore may have been formed by base substitutions in a cryptically simple sequence, followed by expansion of the repeat arrays of certain motifs. The other seven microsatellites appear to be equally likely to have formed by base substitutions or short duplications. Duplications of a small number of bases that created a short repetitive sequence were frequently seen in the regions flanking the microsatellite (Fig. 3).

Degradation of the Nine Microsatellite Loci

The microsatellite sequences mapped to the TEF gene tree (Fig. 1) were also examined to determine if the repetitive sequence had been degraded by loss of repeat units or by accumulation of interrupting point mutations within the repeat array. Of the nine microsatellites, only FoFE5 and FoAB11 appear to have lost their repetitive structure by the removal of repeat units. In locus FoFE5, only one repeat unit is present in isolates *F. o.* 4287, *F. o.* NRRL 26406, and *F. o.* MX4. For locus FoAB11, the microsatellite $(CACAGCA)_n$ in *F. redolens* and *F. oxysporum* appears to have mutated to CACAGCA CAGTA in *F. verticillioides* due to the deletion of CA.

Interruptions of the repeat array were a more common cause of degradation for these loci. Numerous base substitutions interrupting the repeats were observed within the *F. oxysporum* isolates, in loci FoAB11, FoAG11, FoDD7, FoDE7, FoFA4, and FoFE5 (Fig. 1). Independent base substitutions occurred at least twice in FoAB11, and at least eight times in FoAG11. In some cases, the interruption did not completely disrupt the microsatellite, as in isolate *F. o.* Fo47 in locus FoDD7, but in other instances, the microsatellite was completely degraded by a base substitution, as in isolate *F. o.* 4287 in locus FoAB11 (Fig. 1).

Removals of interruptions that restore the perfect repeat array may have occurred two times independently in locus

Table 1 Gene diversity of the microsatellite repeats and regions flanking the microsatellites

Locus	Repeat motif[a]	Microsatellite diversity[b]	Flanking region diversity[c]	Best mutational model	Location
FoAB11	$(CACAGCA)_n$	0.74	0.00945	Multi-step*	Non-coding region
FoAD12	$(AG)_n$	0.78	0.01080	Stepwise*	Non-coding region
FoAG11	$(CT)_n$ $(GT)_n$	0.83	0.03568	Multi-step*	Intron
FoDC5	$(TG)_n$	0.88	0.00291	Stepwise*	Non-coding region
FoDD7	$(CTT)_n$	0.41	0.01581	Multi-step*	Non-coding region
FoDE7	$(CGTT)_n$ CCTT CGTC	0.15	0.00798	Stepwise	3′ UTR
DoDF7	$(GT)_n$	0.28	0.00532	Stepwise	Intron
FoFA4	$(AG)_n$	0.48	0.00415	Stepwise	Non-coding region
FoFE5	$(GTT)_n$	0.15	0.00241	Stepwise	Exon

[a] Repeat motif found in the majority of *F. oxysporum* isolates

[b] Microsatellite diversity $= 1 - \sum x_i^2$, where x_i is the frequency of the ith allele (Nei 1987), based on the size of the core repeat region

[c] Sequence diversity is based on the average pair-wise differences per base pair

* Stepwise and multi-step mutational models were significantly different, $P < 0.05$, log likelihood test

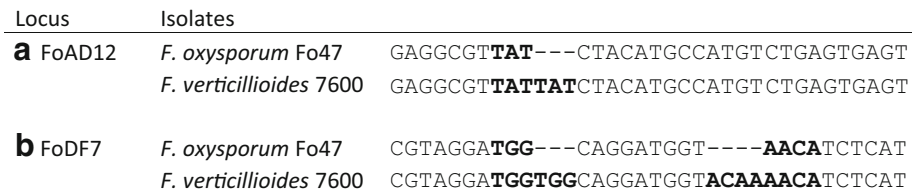

Fig. 3 Examples of insertions/deletions in the microsatellite flanking regions creating cryptically simple sequence. Duplicated regions are shown in bold. (**a**) and (**b**) Examples of duplications, and reversed duplicated sequence in (**b**). **c** Example of a duplication degrading a microsatellite region with the repeat motif AGACAG (*underlined*) but creating the cryptically simple sequence $(GAG)_2$

FoAG11. In a clade where the majority of taxa contain the interrupted microsatellite $(CT)_{7–8}(GT)_{3–4}AT(GT)_4$, five isolates contain the perfect microsatellite $(CT)_{8,10}(GT)_{7,8}$. The AT interruption is found in two taxa outside of the clade, suggesting that it is ancestral and has been removed in the isolates with perfect repeats. An interruption may have also been removed in isolate *F. oxysporum* MN25, which has the perfect repeat $(CT)_{20}(GT)_{10}$, lacking the ATTT interruption found in all other taxa in that clade (Fig. 1).

Utility of Flanking Region Polymorphisms as Genetic Markers

The number of SNPs and alignment gaps between *F. oxysporum* Fo47 and *F. verticillioides* 7600, for which well-assembled genomes are available, were compared to determine if the flanking regions had higher rates of polymorphisms than would be expected. For the flanking regions, there were 12 gaps in the 1,766 bp sequence from *F. oxysporum* Fo47 and seven gaps in the 1766 bp sequence from *F. verticillioides* 7600, for a rate of 6.8 gaps per kb and 3.9 gaps per kb, respectively. Alignment and comparison of the two genomes found 175,169 gaps in the 49,439,464 bp *F. oxysporum* Fo47 genome, for a rate of 4.2 gaps per kb, and 143,810 gaps in the 41,700,345 bp *F. verticillioides* 7600 genome, for a rate of 3.4 gaps per kb, not appreciably different from the rates in the flanking regions. Rates of SNPs were likewise similar between the flanking regions and the entire genomes, with 123 SNPs found in the flanking regions for a rate of 69.6 SNPs per kb for *F. oxysporum* Fo47 and 68.6 SNPs per kb for *F. verticillioides* 7600 and 3940,262 SNPs between the two genomes for a rate of 79.7 SNPs per kb for *F. oxysporum* Fo47 and 94.5 SNPs per kb for *F. verticillioides* 7600.

A phylogeny was inferred based on SNPs in the flanking regions and compared to a phylogeny based on the TEF coding region, commonly used for *Fusarium* phylogenetics. The two phylogenies were similar, with differences only in poorly supported clades (Fig. 4).

Discussion

This study addressed microsatellite evolution in the fungus *F. oxysporum* compared to other eukaryotes. Fungi generally have lower genome densities of microsatellites and

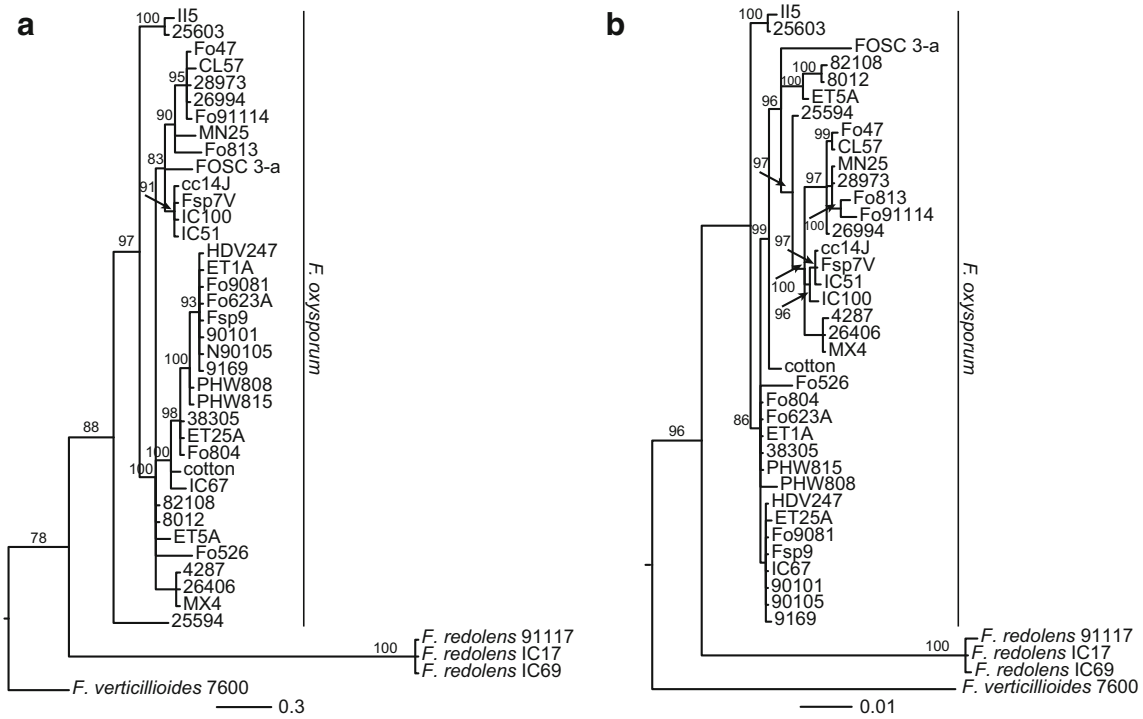

Fig. 4 Comparison of (**a**) a Bayesian tree of the translation elongation factor 1-α (TEF) gene with (**b**) Bayesian tree of the non-repetitive regions flanking the microsatellites. Posterior probability values >75 % are given for each branch for trees in (**a**) and (**b**), indicating the probability that the branches are correct given the phylogenetic model

microsatellites with fewer repeat units than other eukaryotes (Karaoglu et al. 2005; Katti et al. 2001; Tóth et al. 2000), although the reasons for this discrepancy are not known. Nine microsatellites were observed across a diverse sample of *F. oxysporum* isolates to track the birth and death of these loci in fungi. The microsatellite life cycle observed for these loci in *F. oxysporum* appears to be similar to that in other eukaryotes, including the model organism *S. cerevisiae*, the fungus best studied for microsatellites. No evidence was found to suggest that the shorter, less polymorphic microsatellites in fungi is due to different mechanisms of microsatellite formation, mutation, and degradation.

Mutation Rate

The rate of microsatellite mutation has been linked to many factors, including the local mutation and mismatch repair rate and the microsatellite's motif size, number of repeats, and genome location (Sawaya et al. 2012). Of the loci analyzed here, the two microsatellites located in conserved regions, FoDE7 in a 3′UTR and FoFE5 in an exon, had the lowest levels of diversity (Table 1), as would be expected (Sawaya et al. 2012). FoFE5 contains a trinucleotide repeat, as do many microsatellites in exons, most likely because slippage mutations would not introduce frameshift errors (Richard et al. 2008). The compound microsatellite FoAG11, located in an intron, was the most diverse in terms of repeat unit alleles and also contained the highest number of base substitutions within the repeats, suggesting that it may be located in a hypervariable region. The microsatellites with more repeat units generally were more diverse, as has been frequently observed for many microsatellites (Kelkar et al. 2008). For example, locus FoAD12, which contained 4–17 AG repeats, was more diverse than locus FoFA4, which contained 2 or 4 AG repeats, and locus FoDC5, which contained 4–19 TG repeats, was more diverse than locus FoDF7, which contained 4–7 GT repeats. Motif size did not appear to have an effect on locus diversity, as the four most diverse loci were the locus with the longest motif, FoAB11, which contained a seven nucleotide repeat motif, and three dinucleotide microsatellites (loci FoAD12, FoAG11, and FoDC5). Motif size has previously been shown to be negatively correlated to mutability, with microsatellites with longer motifs being less mutable, but mutability is highly variable for loci with the same motif size (Kelkar et al. 2008), so it is unsurprising that a relationship between size and diversity was not seen for the nine loci in this study. Although the overall mutation rate cannot be accurately determined from this dataset, the patterns of mutation for these loci are the same as has been observed in other eukaryotes (Chambers and MacAvoy 2000).

The majority of sequences in the *F. oxysporum* isolates differed from close relatives by only a few units, suggesting that most mutations were stepwise mutations, with a few possible multi-step mutations. This observation fits well with models of microsatellite evolution that assume the gain or loss of one repeat unit at a time, plus some less frequent larger mutations (Henderson and Petes 1992). A majority of microsatellites in other organisms also fit this stepwise model, with some exceptions (Putman and Carbone 2014). For example, one study in humans found almost all tetranucleotide motif microsatellites fit the stepwise model, while 32 % of mutations to dinucleotide motif microsatellites were multi-step (Sun et al. 2012). Exceptions to the stepwise mutational model are well-known confounders in population genetic studies (Putman and Carbone 2014) and are likely to be as much of a problem in studies of fungi as in other organisms.

Microsatellite Life Cycle

Microsatellites are hypothesized to originate from cryptically simple regions by base substitutions, base duplications, insertions, or deletions that mutate the region into a perfect repeat (Kelkar et al. 2011; Zhu et al. 2000b). Kelkar et al. (2011) found that microsatellites in the human genome were more likely to be created by base substitutions for loci with fewer repeats and more likely to be created by insertions and deletions (indels) for loci with more repeats. Not enough loci were studied here to determine if the same pattern holds true in *F. oxysporum*. However, the two compound microsatellites studied appear to have been derived from an ancestral microsatellite that became interrupted by base substitutions, followed by expansion of the new repeat motif, as hypothesized by Kofler et al. (2008).

Twenty-three indels were observed in the flanking regions of *F. oxysporum* microsatellites compared to *F. verticillioides*, as has been observed in the flanking regions of microsatellites for many organisms (Barthe et al. 2012; Dettman and Taylor 2004; Zhu et al. 2000a). Eleven of the indels involving multiple bases in the flanking regions created cryptically simple sequences (Fig. 4), demonstrating how microsatellites may form by insertion/deletion-like events, which appear to be the main mutational mechanisms for adding or subtracting repeats to shorter repetitive elements in cryptically simple sequences (Dieringer and Schlötterer 2003; Kelkar et al. 2010). However, similar numbers of indels per kilobase were calculated for the conserved regions of the *F. oxysporum* and *F. verticillioides* genomes as whole as for the regions flanking the microsatellites, suggesting that the number of indels observed close to these microsatellites is not exceptional.

Microsatellites are hypothesized to "die," or lose their repetitive structure, primarily by the accumulation of point mutations within the repeat array, or occasionally by the loss of all repeat units (Kelkar et al. 2011). In these microsatellites, four loci lost their repetitive structure due to the accumulation of interrupting point mutations and two lost all repeat units. Microsatellites in *F. redolens* and *F. verticillioides* generally had more interruptions than in *F. oxysporum*, most likely due to ascertainment bias, in which long, perfect microsatellites are preferentially selected in the taxon used for primer development, in this case *F. oxysporum* (Brandström and Ellegren 2008). It cannot be determined from these data if the interruptions observed will prevent replication slippage from taking place, as has been previously observed (Ananda et al. 2014; Richard et al. 2008). Base substitutions do not necessarily lead to the death of the microsatellite, as Harr et al. (2000) found that interruptions can be purified from the repeat array, most likely by replication slippage. Evidence suggesting the independent removal of two interruptions was found in loci FoAG11.

Minisatellites

Minisatellites are regions of 10–15 bp long tandem repeats that, despite apparent similarities to microsatellites, can behave much differently than microsatellites. Minisatellites often contain variable repeat regions within each array, may mutate faster during meiosis than mitosis, and may display mutational polarity, meaning that mutations preferentially occur at one end (Richard et al. 2008). Mutations can occur both through replication slippage, similar to microsatellites, and through homologous recombination, especially gene conversion between different alleles, as can be seen by the rearrangement of variable repeats (Richard et al. 2008). The minisatellite described in this study, locus FoED2, appears to be located in the long terminal inverted repeat of a transposon (http://www.broadinstitute.org), which has also been reported for the *F. oxysporum* transposon *marsu* (Hua-Van et al. 2000). Mapping the sequences of the minisatellite FoED2 to the *F. oxysporum* phylogeny showed gains and loss of one or two repeat units at a time and no complex rearrangements of units (Fig. 2), suggesting that mutation by replication slippage rather than gene conversion could have caused all of the observed alleles. *F. oxysporum* has never been observed to produce diploid, sexual structures (Baayen et al. 2001), although the possibility of cryptic sex cannot be excluded (Yun et al. 2000), which may explain the lack of evidence of gene conversion at this allele. Likewise, a minisatellite found in the fungus *Botrytis cinerea* also appears to mutate by replication slippage (Giraud et al. 1998). Because some minisatellites have been found mutate faster during meiosis than mitosis, it seems possible that minisatellites in asexual fungi could evolve differently than in other organisms, but

the number of minisatellites in fungal genomes (Bally et al. 2010) and their contribution to genomic architecture (Richard and Dujon 2006) have only begun to be studied.

Utility of Microsatellite Flanking Regions in Phylogenetics

Flanking region sequences may be highly variable and informative as genetic markers (Barthe et al. 2012; Payseur and Jing 2009). Some studies have suggested that regions flanking microsatellites tend to be more polymorphic than other regions (Brohede and Ellegren 1999; Santibáñez-Koref et al. 2001). For the nine loci in this study, the flanking regions of the microsatellites in this study were not more polymorphic than the average rate of the genome. The observed higher polymorphism rate around microsatellites has been suggested to be due mutations in the microsatellite affecting the surrounding region (Brohede and Ellegren 1999; Santibáñez-Koref et al. 2001), which raises the possibility that these polymorphisms may violate some of the assumptions in phylogenetic reconstruction. The flanking sequences here were polymorphic enough to construct a phylogenetic tree with many well-supported branches, and the tree was similar to the TEF tree in all of the clades with high support (Fig. 4). Flanking region polymorphisms therefore may be useful genetic markers, even if flanking regions may not be more polymorphic than other genomic regions. Researchers designing microsatellite markers should also look at flanking region variation as a potential source of informative characters.

Acknowledgments This work was part of the dissertation of J. E. D., which was funded by the Pennsylvania State University Graduate Fellowship program and a Microbial Functional Genomics Training Fellowship funded to Dr. Seogchan Kang.

References

Ananda G, Hile SE, Breski A, Wang Y, Kelkar Y, Makova KD, Eckert KA (2014) Microsatellite interruptions stabilize primate genomes and exist as population-specific single nucleotide polymorphisms within individual human genomes. PLoS Genet 10:e1004498

Baayen RP, O'Donnell K, Breeuwsma S, Geiser DM, Waalwijk C (2001) Molecular relationships of fungi within the *Fusarium redolens-F. hostae* clade. Phytopathology 91:1037–1044

Bally P, Grandaubert J, Rouxel T, Balesdent MH (2010) FONZIE: an optimized pipeline for minisatellite marker discovery and primer design from large sequence data sets. BMC Res Notes 3:322

Barthe S, Gugerli F, Barkley NA, Maggia L, Cardi C, Scotti I (2012) Always look on both sides: phylogenetic information conveyed by simple sequence repeat allele sequences. PLoS One 7:e40699

Brandström M, Ellegren H (2008) Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. Genome Res 18:881–887

Brohede J, Ellegren H (1999) Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. Proc R Soc Lond B 266:825–833

Buschiazzo E, Gemmell NJ (2006) The rise, fall and renaissance of microsatellites in eukaryotic genomes. BioEssays 28:1040–1050

Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. Compar Biochem Physiol B 126:455–476

Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403

Demers JE, Garzón CD, Jiménez-Gasco MM (2014) Striking genetic similarity between races of *Fusarium oxysporum* f. sp. *ciceris* confirms a monophyletic origin and clonal evolution of the chickpea vascular wilt pathogen. Eur J Plant Path 139:303–318

Dettman JR, Taylor JW (2004) Mutation and evolution of microsatellite loci in Neurospora. Genetics 168:1231–1248

Dieringer D, Schlötterer C (2003) Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. Genome Res 13:2242–2251

Dutech C, Enjalbert J, Fournier E, Delmotte F, Barrès B, Carlier J, Tharreau D, Girand T (2007) Challenges of microsatellite isolation in fungi. Fungal Genet Biol 44:933–949

Fusarium Comparative Database (2012) Broad Institute. Available from http://www.broadinstitute.org/annotation/genome/fusarium_graminearum/MultiHome.html. Accessed Jan 2012

Giraud T, Fortini D, Levis C, Brygoo Y (1998) The minisatellite MSB1, in the fungus *Botrytis cinerea*, probably mutates by slippage. Mol Biol Evol 15:1524–1531

Harr B, Zangerl B, Schlötterer C (2000) Removal of microsatellite interruptions by DNA replication slippage: phylogenetic evidence from *Drosophila*. Mol Biol Evol 17:1001–1009

Henderson ST, Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol Cell Biol 12:2749–2757

Hey J, Wakeley J (1997) A coalescent estimator of the population recombination rate. Genetics 145:833–846

Hua-Van A, Davière JM, Kaper F, Langin T, Daboussi MJ (2000) Genome organization in Fusarium oxysporum: clusters of class II transposons. Curr Genet 37:339–347

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Karaoglu H, Lee CMY, Meyer W (2005) Survey of simple sequence repeats in completed fungal genomes. Mol Biol Evol 22:639–649

Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol 18:1161–1167

Kayser M, Roewer L, Hedman M et al (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Human Genet 66:1580–1588

Kelkar YD, Tyekucheva S, Chiaromonte F, Makova KD (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome Res 18:30–38

Kelkar YD, Strubczewski N, Hile SE, Chiaromonte F, Eckert KA, Makova KD (2010) What is a microsatellite: a computational and experimental definition based on repeat mutational behavior at A/T and GT/AC repeats. Genome Biol Evol 2:620–635

Kelkar YD, Eckert KA, Chiaromonte F, Makova KD (2011) A matter of life or death: how microsatellites emerge in and vanish from the human genome. Genome Res 21:2038–2048

Kofler R, Schlötterer C, Luschützky E, Lelley T (2008) Survey of microsatellite clustering in eight fully sequenced species sheds light on the origin of compound microsatellites. BMC Genom 9:612

Labbé J, Murat C, Morin E, Le Tacon F, Martin F (2011) Survey and analysis of simple sequence repeats in the *Laccaria bicolor*

genome, with development of microsatellite markers. Curr Genet 57:75–88

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Liu M, Rodrigue N, Kolmer J (2014) Population divergence in the wheat leaf rust fungus *Puccinia triticina* is correlated with wheat evolution. Heredity 112:443–453

Ma L-J, van der Does HC, Borkovich KA et al (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. Nature 464:367–373

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nielsen R (1997) A maximum likelihood approach to population samples of microsatellite alleles. Genetics 146:711–716

O'Donnell K, Kistler HC, Cigelnik E, Ploetz RC (1998) Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. Proc Natl Acad Sci USA 95:2044–2049

Paun O, Hörandl E (2006) Evolution of hypervariable microsatellites in apomictic polyploid lineages of *Ranunculus carpaticola*: directional bias at dinucleotide loci. Genetics 174:387–398

Payseur BA, Jing P (2009) A genomewide comparison of population structure at STRPs and nearby SNPs in humans. Mol Biol Evol 26:1369–1377

Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecol Evol 4:4399–4428

Richard G-F, Dujon B (2006) Molecular evolution of minisatellites in hemiascomycetous yeasts. Mol Biol Evol 23:189–202

Richard G-F, Kerrest A, Dujon B (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol Mol Biol Rev 72:686–727

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Santibáñez-Koref MF, Gangeswaran R, Hancock JM (2001) A relationship between lengths of microsatellites and nearby substitution rates in mammalian genomes. Mol Biol Evol 18:2119–2123

Sawaya SM, Lennon D, Buschiazzo E, Gemmell N, Minin VN (2012) Measuring microsatellite conservation in mammalian evolution with a phylogenetic birth–death model. Genome Biol Evol 4:748–759

Schoebel CN, Brodbeck S, Buehler D, Cornejo C, Gajurel J, Hartikainen H, Keller D, Leys M, Říčanová Š, Segelbacher G, Werth S, Csencsics D (2013) Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing. J Evol Biol 26:600–611

Skovgaard K, Rosendahl S, O'Donnell K, Nirenberg HI (2003) *Fusarium commune* is a new species identified by morphological and molecular phylogenetic data. Mycologia 95:630–636

Sun JX, Helgason A, Masson G, Ebenesersdóttir SS, Li H, Mallick S, Gnerre S, Patterson N, Kong A, Reich D, Stefansson K (2012) A direct characterization of human mutation based on microsatellites. Nat Genet 44:1161–1165

Taylor JW, Fisher MC (2003) Fungal multilocus sequence typing—it's not just for bacteria. Curr Opin Microbiol 6:351–356

Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10:967–981

Weetman D, Hauser L, Carvalho GR (2006) Heterogeneous evolution of microsatellites revealed by reconstruction of recent mutation history in an invasive apomictic snail, *Potamopyrgus antipodarum*. Genetica 127:285–293

Yun S-H, Arie T, Kaneko I, Yoder OC, Turgeon BG (2000) Molecular organization of mating type loci in heterothallic, homothallic, and asexual *Gibberella*/*Fusarium* species. Fungal Genet Biol 31:7–20

Zhu Y, Queller DC, Strassmann JE (2000a) A phylogenetic perspective on sequence evolution in microsatellite loci. J Mol Evol 50:324–338

Zhu Y, Strassmann JE, Queller DC (2000b) Insertions, substitutions, and the origin of microsatellites. Genet Res 76:227–236